Towards Knowledge-Based Systems for GDPR Compliance

Harshvardhan J. Pandit, Declan O'Sullivan, and Dave Lewis

ADAPT Centre, Trinity College Dublin, Dublin, Ireland
{harshvardhan.pandit|declan.osullivan|dave.lewis}@adaptcentre.ie

Abstract. Legal compliance is traditionally seen to be sufficiently demonstrable using legal documents that describe how various operations and activities follow a given set of obligations. The General Data Protection Regulation (GDPR) enforces larger responsibilities upon organisations and provides motivation for the use of technological measures that can ease its compliance. While there is no legal requirement to collaborate on compliance technologies or to use a common mechanism for defining knowledge, doing so has several benefits to the larger community. Through this paper, we describe how open and shared technologies targeted towards GDPR and its compliance can be used to create knowledge-based systems. Our approach uses semantic web technologies due to their open and flexible nature towards describing concepts and relationships. We present a model for such a knowledge-based system along with work published to date.

Keywords: GDPR, knowledge-based system, legal compliance graph, legal compliance, provenance, consent, linked data, semantic web

1 Introduction

The General Data Protection Regulation (GDPR) is an European data protection legislation that introduces changes to the way consent and personal data need to be managed by organisations. A large part of the motivation towards efficient adoption of the regulation is the significant amount of fines that could be levied for non-compliance. In this regard, solutions towards its compliance have seen a large amount of interest in the industrial as well as academic community.

Semantic Web provides a common base of technologies and data representation formats that are both open and expressive. Their adoption can aid in the building of common solutions and foster interoperability by virtue of commonly understood knowledge forms. By using the semantic web to combine compliance related data, it is possible to develop knowledge-based systems that can cater to a large area of compliance tasks based on commonality in requirements. In this paper, we present work done to date towards such a knowledge-based system.

2 Work done to date

We have worked [16] on exploring the information flows between different organisations in the context of the GDPR with the goal of identifying a data model for GDPR-related interoperability. We identified entities and the nature of relationships between them using an analysis of the text of the GDPR to categorise relevant articles based on points of interactions between the information flows. Through this, we identified five information categories, which are provenance, data sharing agreements, consent, certification, and compliance along with the dependencies between them. We also presented an evaluation of the available standards based on maturity and recommendation for representation of identified information categories.

To date, we have developed and published representations for provenance called GDPR Provenvanve Ontology (GDPRov) [15] for describing the provenance of consent and data lifecycles using GDPR terminology. We also have investigated possible approaches towards representations for consent [7] and data sharing agreements called Data Protection Rights Language (DPRL) [8]. GDPRtEXT [14] provides a way to refer and link information related to specific articles, terms, and concepts within the GDPR in a machine-readable manner.

This is a crucial aspect towards our aim in building a knowledge-based system. Information representations for certification and compliance are part of our planned future work.

3 Knowledge-based System for Compliance

We primarily express knowledge in the form of RDF triples expressed using suitable OWL ontologies. It is stored within a triple-store, with querying provided by SPARQL[1]. The knowledge-based system and its infrastructure is depicted in Fig. 1, and is based on the consent and data management model previously published [7]. Depending on the requirements of usage, access control mechanisms [11] can be used to ensure authorised accesses for security purposes.

In the context of GDPR compliance, the knowledge base stores facts, assertions, records, and logs pertaining to tasks associated with the maintenance and demonstration of GDPR obligations. The five categories of information, mentioned previously, are used to categorise the information stored within the knowledge base. The information represented by the categories comes from the following data sources:

Data Subject: The data subject provides consent and personal data, for which the knowledge base would store information related to how the consent and data were acquired, and record subsequent changes to consent. Along with this, the exercising of rights would also be recorded as actions involving the data subject.

Data Controller: The bulk of information in the knowledge base relates to and comes from Data Controllers and Data Processors. This includes information about the various activities associated with consent and personal data such

[1] `https://www.w3.org/TR/sparql11-query/`

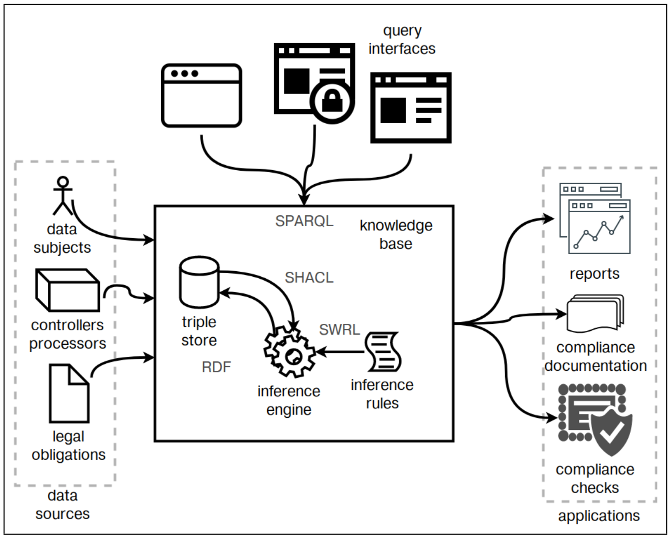

Fig. 1. An overview of the knowledge-based model for GDPR compliance

as collection, storage, sharing, archival, and deletion. This information includes provenance metadata about the activities and how they interact with data, including the provision of various rights and handling data breaches.

Data Processor: A Data Processor acts on the documented instructions of a Data Controller. The knowledge base therefore would contain these instructions in a form that can be queried or combined with other information.

Certification Authority: A Certification Authority awards certifications and seals to organisations based on certain criteria. The criteria and its evaluation mechanisms would be part of the knowledge base for introspection and for demonstration of compliance.

Supervisory Authority: Supervisory Authorities may define additional obligations apart from GDPR towards its compliance. In addition, any communication from or to the Supervisory Authority, such as in the case of data breach, also needs to be stored and maintained for compliance purposes.

Query Interface We envision a web-based interactive interface that allows users to query and explore its results. The interactive aspect of the interface is important as it allows the user to explore more information about the chosen query result. For example, a query for steps that collect consent returns results as a list of items. The user can then click on an item to get more information about that particular step, such as whether it is part of a larger process, or what version of terms and conditions it uses. Having interactive systems allows information to be combined in more dynamic ways, which leads to better interfaces for the underlying knowledge base. The interface would act to allow users to specify

SPARQL queries without knowing the technical complexities of the underlying system.

Inference Engine The quantification of GDPR obligations into inference rules is a complex task. One possibility we intend to explore is the use of SHACL[2], which is a constraint expression language for RDF, to define sets of constraints related to GDPR obligations. This allows the system to check whether the required set of information is present in order for higher-order rules to be executed. For example, using SHACL, it is possible to check whether steps for sharing personal data always have reference to a valid consent or a legal basis as justification. The task of determining whether the sharing itself is compliant with the given consent can then be evaluated using other forms of rule-based inference such as using SWRL[3] with the assumption that all required knowledge exists. This allows inferencing compliance based on constraints while ensuring the data itself is present in the required and correct format.

Linking knowledge using GDPRtEXT The information in the knowledge base coming from different sources would have differing identifiers and may not be related to the required GDPR concepts. Additionally, defining compliance-related information requires a way to uniformly refer to GDPR obligations so that it can be analysed, queried, and retrieved effectively. GDPRtEXT provides a ‘glue’ layer for the linking of related information using GDPR concepts. For example, information related to handling the right to data portability can use the appropriate GDPR terms and concepts to state their relation to this obligation. Queries and results can then retrieve this information to display the intended actions to be taken in the model, the log of what actually happened upon requests, as well as the inference engine’s compliance information using the same concepts and terms as mentioned in the text of the GDPR. In future, we plan to extend the list of terms and concepts, as well as to create additional resources for defining compliance-related terms and concepts specific to the GDPR. The use of GDPRtEXT along with other components of the knowledge-base is provided as an overview in Fig. 2.

4 Applications

The nature of a knowledge-based system changes based on who the intended user(s) are. If the system is targeted towards data subjects, its aim will be to provide information about their personal data and consent, and how it is being collected and used. If the system is developed for controllers and processors, its use will be to assist in the management of compliance information. This involves checking whether the controller or processor fulfils certain obligations such as having systems in place for handling of data breaches and various rights, as

[2] `https://www.w3.org/TR/shacl/`

[3] `https://www.w3.org/Submission/SWRL/`

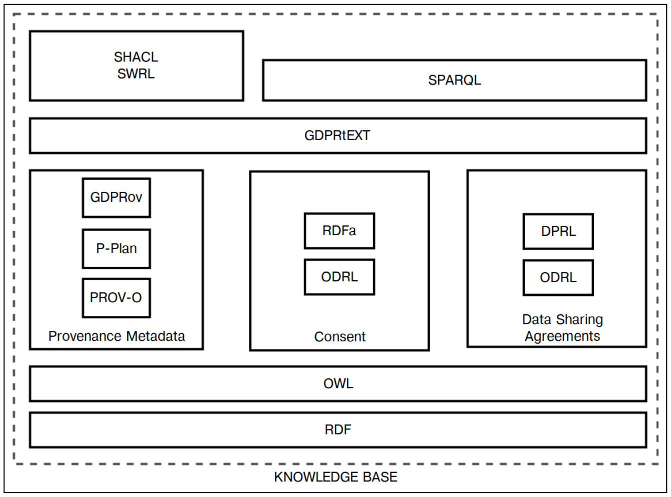

Fig. 2. Semantic Web Technologies used in Knowledge Base

well as providing exploration of the consent and data lifecycles within activities. For instances where privacy or access is a concern, only the metadata can be stored in the system. We describe specific use-cases of such applications below for controllers and processors.

Automated Compliance Checks: Due to the significant amount of potential fines under the GDPR, the maintenance of compliance is an essential activity for organisations. A system that can assist in this process must be scalable to handle a large number of data subjects, which can only be done efficiently through automation of most of its tasks. The knowledge-based system described in this paper provides for such automation through its machine-readable data and query system. Additionally, it is possible to record the entire process and show that due diligence was taken when important changes were made to the system as part of the DPIA process mentioned within the GDPR.

Compliance Documentation: Generation of compliance documentation will be an important activity under the GDPR. Additional information may need to be queried or accessed as part of this process that can sufficiently demonstrate adherence to obligations. For example, showing that personal data is not shared without prior consent can be done by using the abstract model of the system where the activities that share data are shown to depend on the permissions specified within the representation of given consent. Using the knowledge-based system, it is possible to express dynamic queries over the obligations, whose results can be used as a form of compliance documentation. Therefore, the system can help an organisation show adherence to relevant obligations of the GDPR in a comprehensive manner. A periodic review of such documentation by the organisation itself can help in the requirement for periodic assessment of compliance.

5 Related Work

Ontologies An initial work [2] addressed a draft version of the GDPR and presented an OWL2 ontology for data controller duties from GDPR obligations which can be used to structure compliance related information.
Impact Assessment & Visualisation There are existing works that address Data Protection Impact Assessment [3] and Privacy Impact Assessment [17]. Both aim to provide a methodology and a template for assessments in the context of GDPR. There has been work on creating interactive dashboards [4] for data subjects that can show the information flows of their consent and personal data as well as provide features for the handling of various rights. Visualisation has also been applied for representing contracts [6] and legal rules [18]. These are useful as requirements for querying of information within the knowledge base.
Smart Contracts There has been work on developing smart contracts [5] for data sharing agreements between organisations. Such smart contracts can be self-fulfilling and can be automated. The use of Artificial Intelligence techniques [9] has also been explored towards supporting the compliance process.
SPECIAL project The Scalable Policy-aware Linked Data Architecture For Privacy, Transparency and Compliance (SPECIAL) project[4] is an European H2020 project that aims to provide a technical solution involving big-data innovation and privacy-aware data protection. Apart from the publicly available deliverables[5] that describe their findings and reports to date, they have also published their work on building a compliance model for GDPR [1,10]. Our work will be influenced by their approach of modelling consent and compliance as a set of verifiable components, with a focus on query answering.
Knowledge Graphs Building legal knowledge graphs has also seen work in the areas of multilingual services [13]. Such knowledge graphs are expected to assist in the provision of compliance by/through design [12] for them to integrate efficiently into existing legal workflows. An overview of semantic web technologies in the areas of privacy, security, and policies published in the semantic web domain [11] discusses the various problems along with potential solutions and approaches. These are influential for the work discussed in this paper in terms of practical approaches and concerns.

6 Conclusion

GDPR presents motivation and opportunities to apply technological solutions for the compliance of legal obligations. Through this paper, we presented our approach towards building a shared knowledge-based system to assist in compliance related tasks using semantic web technologies. The knowledge-base is based on a GDPR model previously published, and is designed based on the identified information from our work on GDPR interoperability model. The representation of the knowledge is discussed through our work published to date regarding

[4] `https://www.specialprivacy.eu/`
[5] `https://www.specialprivacy.eu/publications/public-deliverables`

metadata representations for provenance, consent, and data sharing agreements in the context of GDPR. The paper discusses the approach towards implementing the presented knowledge-base based on its data sources, inference engine, and usability in a query interface. The paper also discusses potential applications of the knowledge base in automating compliance checks and generating compliance documentation. For future work, we look towards implementing a proof-of-concept knowledge-base from a real-world data to demonstrate the feasibility of the approach.

Acknowledgements

This work is supported by the ADAPT Centre for Digital Content Technology which is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

References

1. Agarwal, S., Steyskal, S., Antunovic, F., Kirrane, S.: Legislative compliance assessment: Framework, model and GDPR instantiation. In: Privacy Technologies and Policy - 6th Annual Privacy Forum, APF 2018, Barcelona, Spain, June 13-14, 2018, Revised Selected Papers. pp. 131–149 (2018). https://doi.org/10.1007/978-3-030-02547-2_8, https://doi.org/10.1007/978-3-030-02547-2_8
2. Bartolini, C., Muthuri, R.: Reconciling Data Protection Rights and Obligations: An Ontology of the Forthcoming EU Regulation. In: Workshop on Language and Semantic Technology for Legal Domain (2015)
3. Bieker, F., Friedewald, M., Hansen, M., Obersteller, H., Rost, M.: A Process for Data Protection Impact Assessment Under the European General Data Protection Regulation. In: Privacy Technologies and Policy. pp. 21–37. Lecture Notes in Computer Science, Springer, Cham (Sep 2016). https://doi.org/10.1007/978-3-319-44760-5_2
4. Bier, C., Kühne, K., Beyerer, J.: PrivacyInsight: The Next Generation Privacy Dashboard. In: Privacy Technologies and Policy. pp. 135–152. Lecture Notes in Computer Science, Springer, Cham (Sep 2016). https://doi.org/10.1007/978-3-319-44760-5_9
5. Corrales, M., Jurcys, P., Kousiouris, G.: Smart Contracts and Smart Disclosure: Coding a GDPR Compliance Framework. SSRN Scholarly Paper ID 3121658, Social Science Research Network, Rochester, NY (Feb 2018), https://papers.ssrn.com/abstract=3121658
6. Esayas, S., Mahler, T., McGillivray, K.: Is a Picture Worth a Thousand Terms? Visualising Contract Terms and Data Protection Requirements for Cloud Computing Users. In: Current Trends in Web Engineering. pp. 39–56. Lecture Notes in Computer Science, Springer, Cham (Jun 2016). https://doi.org/10.1007/978-3-319-46963-8_4
7. Fatema, K., Hadziselimovic, E., Pandit, H.J., Debruyne, C., Lewis, D., O'Sullivan, D.: Compliance through Informed Consent: Semantic Based Consent Permission and Data Management Model. In: Proceedings of the 5th Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn2017) (PrivOn) (2017), http://ceur-ws.org/Vol-1951/#paper-05

8. Hadziselimovic, E., Fatema, K., Pandit, H.J., Lewis, D.: Linked Data Contracts to Support Data Protection and Data Ethics in the Sharing of Scientific Data. In: Proceedings of the First Workshop on Enabling Open Semantic Science (SemSci). pp. 55–62 (2017), `http://ceur-ws.org/Vol-1931/#paper-08`
9. Kingston, J.: Using artificial intelligence to support compliance with the general data protection regulation. Artificial Intelligence and Law **25**(4), 429–443 (Dec 2017). https://doi.org/10.1007/s10506-017-9206-9, `https://link.springer.com/article/10.1007/s10506-017-9206-9`
10. Kirrane, S., Fernández, J.D., Dullaert, W., Milosevic, U., Polleres, A., Bonatti, P.A., Wenning, R., Drozd, O., Raschke, P.: A scalable consent, transparency and compliance architecture. In: The Semantic Web: ESWC 2018 Satellite Events - ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers. pp. 131–136 (2018). https://doi.org/10.1007/978-3-319-98192-5_25, `https://doi.org/10.1007/978-3-319-98192-5_25`
11. Kirrane, S., Villata, S., d'Aquin, M.: Privacy, security and policies: A review of problems and solutions with semantic web technologies. Semantic Web **9**(2), 153–161 (Jan 2018). https://doi.org/10.3233/SW-180289, `https://content.iospress.com/articles/semantic-web/sw289`
12. Mayer, W., Casanovas, P., Stumptner, M.: Semantic Workflows in Law Enforcement Investigations and Legal Requirements. In: Proceedings of the 1st Workshop on Technologies for Regulatory Compliance co-located with the 30th International Conference on Legal Knowledge and Information Systems (JURIX 2017) (2017)
13. Montiel-Ponsoda, E., Rodríguez-Doncel, V., Gracia, J.: Building the Legal Knowledge Graph for Smart Compliance Services in Multilingual Europe. In: Proceedings of the 1st Workshop on Technologies for Regulatory Compliance co-located with the 30th International Conference on Legal Knowledge and Information Systems (JURIX 2017) (2017)
14. Pandit, H.J., Fatema, K., O'Sullivan, D., Lewis, D.: GDPRtEXT - GDPR as a Linked Data Resource. In: 15th European Semantic Web Conference (in-press. Heraklion, Crete, Greece (2018), `http://purl.org/ADAPT/pub/E18ESWC_GDPRtEXT`
15. Pandit, H.J., Lewis, D.: Modelling Provenance for GDPR Compliance using Linked Open Data Vocabularies. In: Proceedings of the 5th Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn2017) (PrivOn) (2017), `http://ceur-ws.org/Vol-1951/#paper-06`
16. Pandit, H.J., O'Sullivan, D., Lewis, D.: GDPR Data Interoperability Model. In: 23 rd EURAS Annual Standardisation Conference (in-press). Dublin, Ireland (2018), `http://purl.org/ADAPT/pub/E18EURAS`
17. Reuben, J., Martucci, L.A., Fischer-Hübner, S., Packer, H.S., Hedbom, H., Moreau, L.: Privacy Impact Assessment Template for Provenance. In: Availability, Reliability and Security (ARES), 2016 11th International Conference on. pp. 653–660. IEEE (2016)
18. Seppala, S., Ceci, M., Huang, H., O'Brien, L., Butler, T.: SmaRT Visualisation of Legal Rules for Compliance. In: Proceedings of the 1st Workshop on Technologies for Regulatory Compliance co-located with the 30th International Conference on Legal Knowledge and Information Systems (JURIX 2017) (2017)